Puzzle

\begin{align*}
\text{Puzzle} & = +1 \\
\text{Puzzle} & = -1 \\
\text{Puzzle} & = ???
\end{align*}
An Answer

\[
\begin{align*}
\text{if symmetric} & : h(x) = +1 \\
\text{otherwise} & : h(x) = -1
\end{align*}
\]

\[ h(x) = \begin{cases} +1 & \text{if symmetric} \\ -1 & \text{otherwise} \end{cases} \]

\[ h\left( \begin{array}{ccc} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{array} \right) = +1 \]
An Answer

\[ h(x) = \begin{cases} 
+1 & \text{if top left is white} \\
-1 & \text{otherwise} 
\end{cases} \]

\[ h \left( \begin{array}{cc}
1 & 0 \\
0 & 0 
\end{array} \right) = -1 \]
An Answer

\[
h(x) = \begin{cases} 
+1 & \text{if } x \in \{\begin{array}{ccc}
1 & 0 & 1 \\
0 & 1 & 0 \\
1 & 0 & 1 \\
\end{array}\} \\
-1 & \text{otherwise}
\end{cases}
\]

\[
h\left(\begin{array}{ccc}
1 & 0 & 1 \\
0 & 1 & 0 \\
1 & 0 & 1 \\
\end{array}\right) = -1
\]
An Answer

\[
h(x) = \begin{cases} 
+1 & \text{if } x \in \{\text{pattern}1, \text{pattern2}, \text{pattern3}\} \\
-1 & \text{otherwise}
\end{cases}
\]

\[
h(\begin{array}{c}
\text{pattern}1 \\
\text{pattern2} \\
\text{pattern3}
\end{array}) = +1
\]
Recall

**Unknown** target function

\[ f: \mathcal{X} \to \mathcal{Y} \]

Training data

\[ \mathcal{D} = \{(x_1, y_1), \ldots, (x_n, y_n)\} \]

Learning Algorithm \( \mathcal{A} \)

Hypothesis Set \( \mathcal{H} \)

Learned Hypothesis \( g: \mathcal{X} \to \mathcal{Y} \)
Analogy

\[
\text{Analogy} = \text{fraction of red marbles in bin}
\]

\[
\mu = \text{fraction of red marbles in bin}
\]

\[
\nu = \text{fraction of red marbles in sample}
\]

Does \( \nu = \mu \)?

Does \( \nu \) say anything about \( \mu \)?
Hoeffding’s Inequality

\[ P\{|\nu - \mu| > \epsilon\} \leq 2e^{-2\epsilon^2 n} \]

- \( \mu = \) fraction of red marbles in bin
- \( \nu = \) fraction of red marbles in a sample of size \( n \)
- As \( n \) increases, RHS decreases
- As \( \epsilon \) decreases, RHS increases
Connection to Learning

- = input space ($\mathcal{X}$)

○ = point in the input space ($\tilde{x}$)

●●●●●● = training data ($\mathcal{D}$)

○ = a point classified correctly by a specified hypothesis $h$

● = a point classified incorrectly by a specified hypothesis $h$

$\nu = \text{fraction of training data classified incorrectly by } h \ (E_{in}(h))$

$\mu = \text{fraction of points in the input space classified incorrectly by } h \ (E_{out}(h))$

$$P\{|E_{in}(h) - E_{out}(h)| > \epsilon\} \leq 2e^{-2\epsilon^2 n}$$
Formal Setup

Unknown target function
\[ f: \mathcal{X} \rightarrow \mathcal{Y} \]

Hypothesis Set
\[ \mathcal{H} \]

Probability Distribution
\[ \mathcal{P} \text{ on } \mathcal{X} \]

Training data
\[ \mathcal{D} = \{(x_1, y_1), \ldots, (x_n, y_n)\} \]

Learning Algorithm
\[ \mathcal{A} \]

Learned Hypothesis
\[ g: \mathcal{X} \rightarrow \mathcal{Y} \]
Validation

- $\mathcal{X}$ = input space
- $\mathcal{D}$ = training data
- $\nu$ = fraction of training data classified incorrectly by a specified hypothesis $h$ ($E_{in}(h)$)
- $\mu$ = fraction of all possible data classified incorrectly by $h$ ($E_{out}(h)$)

$$P\{|E_{in}(h) - E_{out}(h)| > \epsilon\} \leq 2e^{-2\epsilon^2 n}$$
Formal Setup

Unknown target function

\[ f : \mathcal{X} \rightarrow \mathcal{Y} \]

Hypothesis Set \( \mathcal{H} \)

Learned Hypothesis

\[ g = \arg\min_{h \in \mathcal{H}} L(h) \]

Training data

\[ \mathcal{D} = \{(x_1, y_1), \ldots, (x_n, y_n)\} \]

Probability Distribution \( \mathcal{P} \) on \( \mathcal{X} \)

Learning Algorithm \( \mathcal{A} \)
• If you toss a fair coin 20 times, the probability that it comes up heads 20 times is $2^{-20} \approx 1e-6$

• If you toss $2^{20}$ fair coins 20 times each, the probability that at least one coin comes up heads 20 times is $\approx 1 - \frac{1}{e} \approx 0.63$
• Take any coin that came up all heads and apply Hoeffding’s inequality

\[ P\{|E_{in}(g) - E_{out}(g)| > \epsilon\} \leq 2e^{-2\epsilon^2 n} \]

\[ P\{|\text{Fraction of tails in } 20\text{ trials} - \\text{Fraction of tails in } \infty\text{ trials}| > \epsilon\} \leq 2e^{-40\epsilon^2} \]

\[ P\{|0 - P\{\text{this coin coming up tails}\}| > \epsilon\} \leq 2e^{-40\epsilon^2} \]

\[ P\left\{P\{\text{this coin coming up tails}\} > \frac{1}{4}\right\} \leq 2e^{-2.5} \approx 0.15 \]
Suppose $\mathcal{H}$ is finite i.e. $\mathcal{H} = \{h_1, ..., h_m\}$

$$P\{|E_{in}(g) - E_{out}(g)| > \epsilon\}$$

$$\leq P\left\{ \bigcup_{j=1}^{m} |E_{in}(h_j) - E_{out}(h_j)| > \epsilon \right\}$$

$$\leq \sum_{j=1}^{m} P\{|E_{in}(h_j) - E_{out}(h_j)| > \epsilon\}$$

$$\leq \sum_{j=1}^{m} 2e^{-2\epsilon^2 n} = 2(m)e^{-2\epsilon^2 n}$$
Hoeffding’s Inequality (Corrected)

- Suppose $\mathcal{H}$ is finite i.e. $\mathcal{H} = \{h_1, \ldots, h_m\}$
- $E_{in}(g) =$ in-sample error of best hypothesis in $\mathcal{H}$
- $E_{out}(g) =$ out-of-sample error of best hypothesis in $\mathcal{H}$

- $P\{|E_{in}(g) - E_{out}(g)| > \epsilon\} \leq 2(m)e^{-2\epsilon^2n}$

- As $n$ increases, RHS decreases
- As $\epsilon$ decreases, RHS increases
- As $m$ increases, RHS increases